

Content analysis of cancer blog posts*

Sujin Kim, PhD

See end of article for author's affiliation.

DOI: 10.3163/1536-5050.97.4.009

Objectives: The efficacy of user-defined subject tagging and software-generated subject tagging for describing and organizing cancer blog contents was explored.

Methods: The Technorati search engine was used to search the blogosphere for cancer blog postings generated during a two-month period. Postings were mined for relevant subject concepts, and blogger-defined tags and Text Analysis Portal for Research (TAPoR) software-defined tags were generated for each message. Descriptive data were collected, and the blogger-defined tags were compared with software-generated tags. Three standard vocabularies (Opinion Templates, Basic Resource, and Medical Subject Headings [MeSH] Resource) were used to assign subject terms to the blogs, with results compared for efficacy in information retrieval.

Results: Descriptive data showed that most of the studied cancer blogs (80%) contained fewer than 500 words each. The numbers of blogger-defined tags per posting ($M=4.49$ per posting) were significantly smaller than the TAPoR keywords ($M=23.55$ per posting). Both blogger-defined subject tags and software-generated subject tags were often overly broad or overly narrow in focus, producing less than effective search results for those seeking to extract information from cancer blogs.

Conclusions: Additional exploration into methods for systematically organizing cancer blog postings is necessary if blogs are to become stable and efficacious information resources for cancer patients, friends, families, or providers.

INTRODUCTION AND BACKGROUND

Internet sites and resources, including cancer blogs, have emerged as popular health communication media among cancer patients, their companions and caregivers, and health care professionals [1, 2]. Cancer patients and their companions or caregivers use blogging to discuss disease issues, share personal stories, and connect with friends and support networks. Uses of cancer blogs also include expanding one's cancer-related knowledge, seeking opinions about or validating information received from health care providers, and preparing information for provider visits [3]. Caregivers or families seek out cancer information such as alternative treatment options, recent news, diagnostic and prognostic information for validation, and emotional relief. They also find information to assist cancer patients who may not have the physiological or psychological strength left for blogging [4, 5]. Not surprisingly, most of the shared medical information in blogs is typically specific to personalized disease matters, sometimes based on the blogger's own experience with the disease. Increasingly, however, health care professionals also use web-based tools like blogs to share clinical knowledge, professional values, and personal experiences [6–10]. Blogging as a writing activity can promote good health. In fact, studies on the healing effect of expressive writing, based on the aspect of online social support, have been reported in the oncology social work and psychology literature [11].

* This publication was made possible by grant number P20RR-16481 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH). Its contents are solely the responsibility of the author and do not necessarily represent the official views of NCRR or NIH.

Highlights

- Cancer patients, families, and providers increasingly use blogs and other communications technologies to learn about health, share information, and support one another.
- User-defined tagging is a popular method for providing subject or topic access to blog messages, but user-defined tags are not centrally defined or based on a standardized vocabulary or thesaurus.
- This study implemented a variety of software-driven text mining techniques to establish descriptive characteristics for 485 collected cancer blogs, as well as blogger-assigned and software-generated tags.

Implications

- Mechanisms for systematically and accurately describing blog messages are necessary to better support bloggers, as well as those seeking relevant content in cancer blogs.
- Automatically generated subject terms such as the Opinion Templates, Basic Resource, or Medical Subject Headings (MeSH) Resource Templates, used in combination with user-defined tags, can provide powerful subject access to blog messages.

Mechanisms for more systematic organization of blog messages, specifically in cancer contexts, are necessary to better support bloggers and those seeking to extract quality information from cancer blogs. Collaborative tagging efforts are increasingly

popular for providing topic- or subject-based access to blog messages; however, tags are not centrally defined or generally based on a standardized thesaurus. User-defined tags, like those used in blogging, were originally introduced as an organizational tool for online contents. These social tagging systems “let their users organize personal resource collections with tags” [12]. Yet, blogger-defined tags are highly personal and inconsistent [4]. Individual bloggers and public blog audiences are not typically trained to organize messages into meaningful categories for later retrieval. When assigning subject tags to blog messages, the average blogger does not use consistent, meaningful thesauri terms. Additionally, most blogs are not semantically organized for retrieval.

Clearly, the Internet and other electronic communication tools are changing the way cancer patients and providers receive and provide information and support. Although the reasons for reading cancer blogs are well documented, how the contents of cancer blogs are organized and structured for information retrieval has not received as much study. This research addressed this issue by collecting, counting, and analyzing the descriptive characteristics and primary subject content of cancer blog messages. Content analysis, text analysis, and data mining techniques were applied to establish descriptive statistics, including the most frequently used words and concepts in blog messages and the number and quality of blogger-defined tags and Text Analysis Portal for Research (TAPoR)-generated subject tags. Blogger-assigned tags and TAPoR-generated tags were compared, as was the efficacy of Opinion Templates, Basic Resource, and Medical Subject Headings (MeSH) Resource for representing blog subject content based on the assignment of subject terms. These results have implications for improving both blogger-generated tagging and thesauri-driven indexing of blogs for increased retrieval and use.

LITERATURE REVIEW: ORGANIZING BLOG MESSAGES

Keeping up with written blog messages can be overwhelming, particularly because most blog users do not have time to visit multiple sites and sort through the necessary information as often as they would perhaps like to do. To improve accessibility to the growing numbers of blog messages, a systematic and consistent means of organizing messages becomes essential. Some of the current work in the area of blogging is intended to develop technologies for organizing and delivering textual information (such as messages posted on blogs) based on their subject content [13–16]. For example, really simple syndication or rich site summary (RSS) web feeds are one common method of accessing and organizing vast amounts of web-based information. Blog readers can select “areas of interest” and then have related RSS feeds delivered directly to them. Some blogging communities, such as LiveJournal, enable bloggers to assign topics of “interest” as messages are posted.

Messages with assigned “interest” tags are then made searchable [15]. However, these “areas of interest” are often too simplistic or generic to systematically retrieve relevant blog messages. Available subject categories such as “types of cancers” (including “breast cancer,” “prostate cancer,” “colorectal cancer,” etc.); “types of therapy” (such as “alternative therapies,” “chemotherapy,” “drug,” and “prevention”); “fundraiser” (“cancer events” and “pink products”); or “celebrities” (including “cancer diagnosis,” “cancer treatment,” and “fundraisers”) are used, but these can be either too broad or too narrow in scope [17]. More importantly, the terms chosen by bloggers and assigned as “areas of interest” are neither consistently nor necessarily adequately representative of message contents.

Collaborative tagging—also known as folksonomy, social classification, social indexing, or social tagging—allows blog authors and readers themselves to assign tags, or subject terms, to blogs [18, 19]. Unlike conventional subject indexing through controlled vocabularies such as MeSH or the Unified Medical Language System (UMLS) or classification schemes such as Library of Congress Classification or Dewey Decimal Classification, advocates of collaborative tagging argue that “allowing meaning of a tag to emerge through collective usage produces a more accurate meaning than if it was defined by a single person or body” [20]. However, unfamiliarity with blog contents and a lack of knowledge of vocabulary and classification control for assigning blog post tags contribute to irrelevant tagging and subsequent poor retrieval of cancer information from blogs [21]. The lack of a standard for blog subject indexing or categorization limits keyword-based access.

Few empirical studies have examined the subject content of cancer blog messages, which is a first step in understanding how best to classify blog messages in predefined categories. Additional research examining the primary subjects of cancer blogs and exploring options for effective indexing and organization of blog postings is important, particularly because blogs are an increasingly accepted form of consumer health and cancer information. This study will help to understand the content of blogs and explore issues related to more consistent and systematic categorization of messages.

METHODOLOGY

This project experimented with a variety of methods for establishing and analyzing descriptive characteristics and relevant subject tags for a set of 485 cancer blog posts. The blog messages were first mined for relevant subject concepts, and then blogger-defined tags were collected and software-defined subject tags were generated for each message. Descriptive data were derived for both blog postings and resulting subject tags. Three standard vocabularies (Opinion Templates, Basic Resource, and MeSH Resource) were used to assign subject terms to the blogs, with results

compared for searching efficacy. The specific research questions addressed in this study were:

1. What are the descriptive characteristics of cancer blog posts?
2. Which specific subject topics are most frequently represented in cancer blog posts, as determined by (a) blogger-defined tags and (b) TAPoR software-generated tags?
3. What are the strengths and weaknesses of sample keyword terms generated by Opinion Templates, Basic Resource, and MeSH Resource Templates?

For the purposes of this study, a "cancer blog post" was defined as a narrative message authored by anyone whose main topic was related to cancer. The author used Technorati <<http://www.technorati.com/search?advanced>>, a web-based blog search engine, to collect the cancer blog posts and their blogger-assigned tags for addressing research questions 1 and 2. Technorati searched the entire blogosphere for relevant entries based on identifying "cancer" as a specific, blogger-defined tag. Other tag words such as "tumor" or "adenocarcinoma" were also searched, but results yielded no more than three postings. The search was limited to the English-language and to blog postings written within the previous few months (September–November 2007). Only initial blog entries or postings were analyzed. All replies, also called "comments," were removed because fewer than 5% of the posts included associated replies. Approximately 1,220 postings were retrieved through this search. Manual selection and validation were performed to remove additional postings irrelevant to cancer as a disease or health-related topic, such as those on the zodiac sign of cancer. After removing irrelevant posts, 485 cancer blog postings were used for analysis.

To address research question 1, TAPoR List Words was used to generate a list and count of the individual words in each blog posting, including unique and duplicate words. Glasgow stop words such as "a," "an," "the," and so forth were excluded from the list. PractiCount and Invoice listed and counted distinct words found in the entire collection of cancer blog posts. For research question 2, SPSS (version 16) measured the occurrences and distribution of words found in the collected postings, and a list of blogger-defined tags was generated, as was a list of TAPoR Keywords Finder-generated subject tags <<http://taporware.mcmaster.ca/~taporware/betaTools/keywordFinder.shtml>>. TAPoR software collected statistics such as the highest frequency, average frequency, and most commonly occurring subject terms. TAPoR Comparator was used to analyze the two sets of subject keywords to identify the most frequently found tags for addressing research questions 1 and 2.

The author further explored the applicability of thesauri indexing to cancer blog postings through research questions 3. The essential subject topics for each posting were identified using SPSS text mining software, Clementine, then subject terms were chosen from each of three English-language Clementine

Resource Templates[†], including Opinion Templates, Basic Resource, and MeSH Resource. Text mining for SPSS Clementine (version 12) was chosen to classify the collected blog posts by automatically generated thesauri because it utilizes the MeSH Resource Template. While the MeSH Resource Template is more specialized to extract medically related concepts specific to MeSH vocabularies, the Basic Resource Template and Opinion Templates were also utilized to extract concepts in a general domain of textual corpus analysis that were not specific to the medical domain. The author reviewed similarities and differences between the generated concept terms, providing useful information that may be considered when developing suggested terminology for blog tagging and retrieval.

RESULTS

Research question 1 sought to describe general characteristics of 485 cancer blogs written between September 1, 2007, and November 31, 2007. Although it is not directly relevant to the subject matter of the blog posting, the average length of a posting is a potentially important factor in estimating the numbers of tags and automatic keywords generated for each posting. The average number of single words found in each posting was 335, ranging from a minimum of 18 words to a maximum of 3,432 words per post. After duplicate word entries were removed, the average number of unique words per posting was 175. Slightly more than 80% ($n=378$) of collected posts contained fewer than 500 words overall. This result showed that the majority of cancer bloggers wrote relatively short postings, consisting of fewer than 3 pages (12-point font, double spaced).

Research question 2 sought to identify the most common topics or primary subject content of the collected postings ($n=485$). Primary subjects were identified based on a count of single word frequencies in the full postings, blogger-assigned tag words, and TAPoR-generated keyword collections. Table 1 displays the top 20 words appearing in any of the collections, ranked by frequency. Of the top 20 words found, 10 words appeared in all 3 word collections; 11 were found in 2 of the collections; and 38 words occurred in just 1 collection. Not surprisingly, "cancer" and "breast" ranked as the first and second most prevalent words in all 3 collections. Because breast cancer is largely a female disease, the topic word "women" also ranked highly in all 3 word lists. This finding supported previous research showing that breast cancer was the most frequently discussed type of cancer online [20, 22, 23].

Descriptive statistics were also collected for the blogger-defined tag and TAPoR-generated keyword collections. Results of word counts in these collections

[†] SPSS Clementine Resource Templates refer to "a set of specialized libraries, which are made up of dictionaries used to define and manage types, terms, synonyms, and exclude lists."

Table 1

Top 20 single words in full posts, blogger-defined tags, and Text Analysis Portal for Research (TAPoR)-generated keyword collections

Rank by frequency	Postings	Frequency	Blogger-defined tags	Frequency	TAPoR-generated keywords	Frequency
1	cancer	2,326	cancer	537	cancer	1,547
2	breast	1,045	breast	234	breast	759
3	pink	413	health	118	awareness	320
4	women	396	awareness	100	pink	267
5	patients	356	pink	48	month	173
6	know	307	month	34	research	134
7	time	303	treatment	29	like	126
8	just	290	prostate	23	new	98
9	new	280	fashion	23	women	93
10	like	275	charity	19	patients	93
11	study	269	medicine	15	health	91
12	said	267	chemotherapy	14	prostate	85
13	treatment	255	genetic	14	american	84
14	awareness	252	issues	14	national	83
15	month	249	family	13	foundation	82
16	risk	248	lung	13	ribbon	78
17	people	230	disease	13	com	75
18	research	223	cure	13	treatment	74
19	cells	209	genetics	13	year	73
20	health	203	care	12	day	63

showed that 1,335 unique words were found in the blogger-defined tag word collection (3,638 total words) and 4,888 unique words (20,500 total words) were found in the TAPoR-generated keyword collection. Slightly more than 36% of the blogger-defined tag words were uniquely identified, while 23.8% of the TAPoR-generated words were uniquely identified. Table 1 illustrates, however, that some frequently occurring words do not actually represent the topic concepts of the cancer blog posts. These nonrepresentative words included "know," "time," "just," "new," "like," "said," "make," and "going" and appeared less frequently in the blogger-generated tag word collection than in the TAPoR-generated collection. These findings indicated that blogger-defined tags used fewer common terms than those of the TAPoR-generated keywords and that TAPoR assigned more irrelevant subject terms than human bloggers did.

Data analysis also identified inconsistencies between the numbers of blogger-assigned tags and the numbers of TAPoR-generated keywords for the same posts, showing that human taggers were more likely to assign an inconsistent number of tags per posting. On average, bloggers assigned 4.49 tags per blog posting and TAPoR generated 23 keywords per posting. The most frequently human-assigned number of tags per post was 4 ($n=80$ posts), followed by 5 tags ($n=62$ posts), followed by 3 tags per post ($n=54$ posts). TAPoR was consistent, with over 65% of the blog posts assigned 23 TAPoR keywords; however, the range of human-assigned tags varied more broadly. Raw number of tags and keywords did not indicate quality of subject indexing, however, because more is not necessarily better.

The twenty most frequently used blogger-defined tags and TAPoR-generated keywords, including those that were hierarchically or contextually related, are ranked by frequency and shown in Table 2. These terms included concepts such as "cancer," "breast cancer," "prostate cancer," "lung cancer," "breast

cancer awareness," "breast cancer research," "breast cancer treatment," "national breast cancer," and "breast cancer foundation." Word variation was an issue observed in the frequently found tags and keywords collection. Such variants included abbreviations (e.g., "dr," "doctor"); plurals ("cell," "cells"); synonyms ("research," "study"); and spelling variations ("breastcancer," "breast+cancer," "breast cancer"). Case sensitivity posed no problem because all words were transformed into lower case to reduce the incidence of word variation. Based on a simple frequency analysis of the collected tags, this study predicted that human tagging may not be an economic method. However, this assumption has a major limitation in that the study could not control the number of keywords generated by TAPoR. These findings are a potential indicator of retrieval performance, with important implications for overall quality of subject accessibility, in terms of breadth and depth of content coverage.

To assess the efficacy of using standardized thesauri terms for indexing and organizing cancer blog contents, SPSS Clementine text mining software was used in correlation with Opinion Templates, Basic Resource Templates, and the MeSH Resource Template to identify core concepts and assign subject terms (concept terms) to each blog post. The resource template-based analyses identified 78 subject concepts in all. Of those 78 concepts, 26 were common to all 3 resource templates. These appear in Table 3. Twenty-three terms were found in 2 of the 3 templates, and 26 terms were uniquely identified in only 1 of the 3 templates used. As shown in Table 3, the term "tumor" was the most frequently found concept in each of the 3 resource templates. Nearly 65% of the MeSH Resource Template posts ($n=315$) used the term "tumor." Nearly 54% ($n=259$) of the postings categorized in Opinion Templates and 52% ($n=252$) of the Basic Resources Template terms also included the word "tumor." These results indicated that the MeSH Resource Template was the best

Table 2

Top 20 blogger-defined tags and TAPoR-generated keywords including multi-words, ranked by frequency

Rank by frequency	Blogger-defined tags	Frequency	TAPoR-generated keywords	Frequency
1	cancer	185	breast cancer	197
2	breast cancer	112	breast cancer awareness	83
3	health	70	cancer awareness month	75
4	breast cancer awareness	58	time	65
5	breast	19	cancer awareness	50
6	charity	18	just	43
7	pink	16	awareness month	42
8	prostate cancer	15	women	38
9	2007 breast cancer awareness month	14	cancer research	37
10	chemotherapy	14	study	36
11	family	13	like	32
12	breast cancer awareness month	10	cancer	31
13	cancer treatment	10	breast cancer research	30
14	awareness	9	people	30
15	health issues	9	way	30
16	lung cancer	8	year	29
17	shopping	8	prostate cancer	27
18	women	8	october	26
10	death	7	support	26
20	fashion	7	day	25

thesaurus of the 3 in consistently identifying “tumor” as a concept word. These findings implied that the selection of the resource template (thesauri) is an important factor in identifying key concepts. Furthermore, retrieval performances can be increased by refining the templates.

In each of the three templates, “key concepts” were linked to more focused “descriptor” terms, but specific descriptor terms varied by template. For example, the descriptor associated with the concept of “patients” in Opinion Templates differed from those in Basic Resource Template and the MeSH Resource Template descriptors. In addition, synonymous concepts such as “medicine,” “treatment,” and “pharmaceutical preparation” were only linked in Opinion Templates. This indicated a lack of standard semantics among generated concepts. The descriptors, however, showed strong potential for use as a

source of word variations for further refinement of defined concepts.

The nature of cancer requires patients and their families or caregivers to learn about the illness, make informed decisions about their treatment options, and cope with a stressful disease experience [24, 25]. Subject concepts identified by the resource templates illustrated both cognitive and emotional processing [11], as illustrated through core concepts such as “side effects of treatment,” “impact on family and friends,” “self-care issues,” and “social activities,” all of which concepts Degner et al. have developed [26]. The “emotion” concept, for example, included specific descriptors such as “worry,” “trouble,” “outrage,” “happiness,” “excitement,” “fear,” and “anxiety.” Concepts also involved medically related topics such as “treatment,” “research,” “results,” “pharmaceutical preparation,” “science,” “progress,” “evidence,”

Table 3

Blog post frequencies for the top 20 concepts from SPSS Clementine models using three resource templates

Opinion Templates Resource Template		Basic Resource Template		MeSH Resource Template	
Concept category	Count	Concept category	Count	Concept category	Count
tumor	259	tumor	252	tumor	315
investigation	206	breast	221	health	181
healthy	199	investigation	182	investigation	181
breast cancer	177	evidence	161	breast	175
effect	172	health	137	evidence	158
breast	165	treatment	136	pharmaceutical preparation	143
drugs	145	professionals	134	biology	133
medicine	143	drugs	130	north america	127
woman	136	effect	130	woman	126
treatment	135	north america	127	treatment	125
flowers	127	woman	127	europa	109
evidence	123	patients	105	patients	107
countries	110	flowers	97	emotions	104
patients	110	emotions	95	flowers	100
possible	107	journal	90	study	91
clean	103	study	88	journal	87
journal	102	awareness	81	effect	86
problem	99	writer	80	awareness	82
organs	97	foods	76	kids	78
study	97	kids	75	foods	76

"investigation," and "drug." The presence of emotive, cognitive, and medically related subject concepts indicated that the collected cancer blogs were used for functions including emotional support and sharing of medical information. More rigorous studies comparing categories of cancer information needs with text mining-generated core categories could help assess the "aboutness" of cancer blog posts [26].

Automatically generated keywords in combination with blogger-defined tags could prove to be powerful subject access mechanisms to inadequately described blog postings. The current, user-driven approach to social tagging provides immediate and obvious benefits to the users, because content is annotated based on the personal understanding of its authors through their own natural language. However, user-driven indexing is, in general, less consistent, less clear, less complete, and thus generally less useful for quality and comprehensive retrieval [12]. Despite the weaknesses of user-generated tags, these research findings showed, however, that blogger-defined tags are representative in a way that annotation descriptors of scientific literature are not. For instance, the human-generated tag collection contained useful descriptors not found in the MeSH thesaurus, such as "cancer awareness month," "cholesterol lowering drugs," "smoking-at-home," "anticancer," and so on. This finding was not entirely unexpected because cancer blog messages are individual narratives, far different than more structured research articles. To achieve maximum benefit from tagging, it is highly advisable to examine some of the collective and collaborative tagging practices and patterns, especially in cancer contexts.

DISCUSSION AND CONCLUSIONS

This study implemented a variety of software-driven text mining techniques to establish descriptive characteristics for 485 collected cancer blogs, as well as blogger-assigned tags and software-generated tags for each blog (research question 1). Collected cancer blogs contained topics that varied from personal stories in narrative writing to sharing of medical information. Analysis of the blogger-assigned and TAPoR-generated tags identified the blogs' primary subject matter (research question 2) and evidenced benefits and drawbacks of both blogger-assigned and software-assigned tags. Additionally, this study compared the efficacy of three standard vocabularies (Opinion Templates, Basic Resource, and MeSH Resource) for realistically tagging cancer blogs. Advanced text analysis techniques, such as text mining, were useful for classifying cancer blog posts to generate major subject concepts, and the Clementine core content descriptors were helpful in identifying subject terms. The automated process identified relevant and detailed descriptors, but it also produced descriptors that were fragmented definitions of concepts. This illustrated the continued need for human intervention, although fragments are often helpful for further refining hierarchical relationships among sub-

concepts. For instance, descriptors of the subject "radiation" suggest sub-concepts such as "radiation treatment," "radiation oncologist" "standard treatment," and "dose of radiation."

Limitations of this study included the fact that only Technorati was used to search the blogosphere and a small, two-month sample was collected. Different results might have resulted from a larger sample or multiple blogosphere search engine results. Future research may want to validate findings through manual indexing, but that was outside the scope of this research. In addition, the study did not uncover complex and varying text mining algorithms that could produce different subject descriptors for the same blog collection.

While manual indexing and professional annotations remain the gold standard for subject indexing, the findings of this research demonstrate that automated techniques based on term frequency can facilitate the process and produce meaningful descriptors. Automatically generated keywords—combined with user-defined tags, assigned via a low-impact, easily understood method—would ultimately assist users in more effectively accessing cancer blog posts. These findings can affect cancer blog retrieval, which is a major issue in online information retrieval but has yet to gain significant attention in the scholarly literature.

It has been established that the public turns to the Internet to communicate with one another about health-related issues including cancer [11, 27]. Considering the ever-growing amount of online biomedical information and newly generated data available through alternative media such as blogs and social networking, it is apparent that the automated processing of text analysis for subject annotation and retrieval has strong potential for improving organization of and access to cancer blogs and other alternative media [28–30]. Medical information and emotional support given through expressive writing online are attracting attention as research topics in cancer information production and retrieval. However, peer-production works, social networking, virtual worlds, and social tagging may not transfer seamlessly into traditional medical library settings [31–33]. In the cancer blogging world, the people who author blogs and create tags do not prepare their work for retrieval in the same ways that indexers of conventional surrogate records prepare their information for effective retrieval. At the same time, however, scholarly terminologies controlled by various medical thesauri cannot easily represent the narrative and expressive writing of cancer bloggers.

These findings are important for identifying core subject topics of cancer blogs and potential candidates for subject descriptors for improved retrieval. Recognizing that resources such as blogs are different than mainstream communication channels in terms of topics and subject access methods, medical librarians can assist in identifying more representative tags, particularly as blog contents are integrated into library collections. Medical librarians have the skills

to teach users not only how to search effectively for information, but also how to represent peer-production materials for improved retrieval.

ACKNOWLEDGEMENT

The author is grateful to graduate research assistants at the University of Kentucky who helped me collect tags, run TAPoR software, and proofread the manuscript.

REFERENCES

1. Frisse ME, Braude RM, Florance V, Fuller S. Informatics and medical libraries: changing needs and changing roles. *Acad Med*. 1995 Jan;70(1):30–5.
2. Thielst CB. Weblogs: a communication tool. *J Health Manag*. 2007;52(5):287–9.
3. Narimatsu H. Detailed analysis of visitors to cancer-related web sites. *J Clin Oncol*. 2008;26(25):4219–23.
4. Kim S, Chung DS. Characteristics of cancer blog users. *J Med Libr Assoc*. 2007 Oct;95(4):445–50.
5. Hillan J. Physician use of patient-centered weblogs and online journals. *Clin Med Res*. 2003;1(4):333–5.
6. Chung DS, Kim S. Blogging activity among cancer patients and their companions: uses, gratifications and predictors of outcomes. *J Am Soc Inf Sci Technol*. 2008 Jan;59(2):297–306.
7. Lagu T, Kaufman EJ, Asch DA, Armstrong K. Content of weblogs written by health professionals. *J Gen Intern Med*. 2008;23(10):1642–6.
8. Santoro E. Podcasts, wikis and blogs: the Web 2.0 tools for medical and health education. *Recent Prog Med*. 2007;98(10):484–94.
9. Boulos MN, Maramba I, Wheeler S. Wikis, blogs and podcasts: a new generation of web based tools for virtual collaborative clinical practice and education. *BMC Med Educ*. 2006;6:41.
10. Poonawalla T, Wagner RF Jr. Assessment of a blog as a medium for dermatology education. *Dermatol Online J*. 2006;12(1):5.
11. Sethi SK. Blog/web log—a new easy and interactive website building tool for a non-net savvy radiologist. *J Thorac Imaging*. 2007;22(2):115–9.
12. Good BM, Kawas EA, Wilkinson MD. Bridging the gap between social tagging and semantic annotation: E.D. the Entity Describer [Internet]. [cited 8 Nov 2008]. <<http://proceedings.nature.com/documents/945/version/1/>>.
13. Morgan NP, Graves KD, Poggi EA, Cheson BD. Implementing an expressive writing study in a cancer clinic. *Oncologist*. 2008;13(2):196–204.
14. Gruhl D, Guha R, Liben-Nowell D, Tomkins A. Information diffusion through blogspace. In: *WWW2004: The 13th International World Wide Web Conference*. New York: NY, ACM Press; 2004. p. 491–501.
15. Herring S, Scheidt L, Bonus S, Wright E. Bridging the gap: a genre analysis of weblogs. In: *HICSS 2004: Proceedings of the 37th Annual Hawaii International Conference on System Science*. Washington, DC: IEEE Computer Society. p. 40101.2.
16. Herring SC, Kouper I, Scheidt LA, Wright E. Women and children last: the discursive construction of weblogs [Internet]. 2005 [cited 8 Nov 2008]. <http://blog.lib.umn.edu/blogosphere/women_and_children.html>.
17. Herring SC, Scheidt LA, Bonus S, Wright E. Bridging the gap: a genre analysis of weblogs. *Proceedings of the 37th Annual Hawaii International Conference on System Science*, Track 4. [Internet]. [cited 8 Nov 2008]. <<http://www.ics.uci.edu/~jpd/classes/ics234cw04/herring.pdf>>.
18. LiveJournal [Internet]. [cited 8 Nov 2008]. <<http://www.livejournal.com>>.
19. Quintarelli E. Folksonomies: power to the people [Internet]. Paper presented at: ISKO Italy-UniMIB Meeting [cited 8 Nov 2008]. <<http://www.iskoi.org/doc/folksonomies.htm>>.
20. Brooks CH, Montanez N. An analysis of the effectiveness of tagging in blogs *Proceedings of the 2005 AAAI Spring Symposium on Computational Approaches to Analyzing Weblogs*; Stanford, CA, Mar 2006.
21. Shirky C. Many2many blog: folksonomy [Internet]. [cited 8 Nov 2008]. <<http://www.many.corante.com/archives/2004/08/25/folksonomy.php>>.
22. Fogel J, Albert SM, Schnabel F. Internet use and social support in women with breast cancer. *Health Psychol*. 2002;21:398–404.
23. Hwang JM, Barnett GA. The use of the Internet for health information and social support: a content analysis of online breast cancer and prostate cancer discussion groups. Paper presented at: Association of Internet Researchers Annual Conference; Chicago, IL, 5–9 Oct 2005.
24. Shaw BR, Hawkins R, McTavish F, Pingree S, Gustafson DH. Effects of insightful disclosure within computer mediated support groups on women with breast cancer. *Health Commun*. 2006;19(2):133–42.
25. Ankem K. Use of information sources by cancer patients: results of a systematic review of the research literature. *Inf Res [Internet]*. 2006;11(3) [cited 8 Nov 2008]. <<http://www.informationr.net/ir/11-3/paper254.html>>.
26. Degner LF, Kristjanson LJ, Bowman D. Information needs and decisional preferences in women with breast cancer. *JAMA*. 1997 May 14;277(18):1485–92.
27. Lancaster FW. Indexing and abstracting in theory and practice. Champaign, IL: University of Illinois; 2003. p. 13–8.
28. McCosker A. Blogging illness: recovering in public. *M/C Journal [Internet]*. 11(6) [cited 8 Feb 2009]. <<http://www.journal.media-culture.org.au/index.php/mcjournal/article/viewArticle/104/>>.
29. Skinner H, Biscope S, Poland B. Quality of Internet access: barrier behind Internet use. *Soc Sci Med*. 2003;57:875–80.
30. Rainie L. The state of blogging [Internet]. Pew Internet & American Life Project. [cited 8 Nov 2008]. <http://www.pewinternet.org/~media/Files/Reports/2005/PIP_blogging_data.pdf>.
31. Ziebland S. The importance of being expert: the quest for cancer information on the Internet. *Soc Sci Med*. 2004;59:1783–93.
32. Sostack M. From Web 2.0 to health literacy: the Buzz at MLA 2007. *Inf Today*. 2007;24(7):28.
33. Scola-Streckenbach S. Experience-based information: the role of web-based patient networks in consumer health information services. *J Consumer Health Internet*. 2008;12:3:216–36.

AUTHOR'S AFFILIATION

Sujin Kim, PhD, sujinkim@uky.edu, Assistant Professor, School of Library and Information Science and Department of Pathology and Laboratory Medicine, University of Kentucky, 339 Lucille Little Building, Lexington, KY 40506-0224

Submitted November 2008; accepted May 2009